

⑫ 公開特許公報(A) 平1-124193

⑨ Int. Cl.⁴

G 11 C 11/34

識別記号

3 6 2

庁内整理番号

C-8522-5B

④ 公開 平成1年(1989)5月17日

審査請求 未請求 発明の数 1 (全14頁)

④ 発明の名称 半導体記憶装置

② 特 願 昭62-281619

② 出 願 昭62(1987)11月6日

⑦ 発 明 者 藤 島 一 康 兵庫県伊丹市瑞原4丁目1番地 三菱電機株式会社エル・エス・アイ研究所内

⑦ 発 明 者 松 田 吉 雄 兵庫県伊丹市瑞原4丁目1番地 三菱電機株式会社エル・エス・アイ研究所内

⑦ 発 明 者 朝 倉 幹 雄 兵庫県伊丹市瑞原4丁目1番地 三菱電機株式会社エル・エス・アイ研究所内

① 出 願 人 三菱電機株式会社 東京都千代田区丸の内2丁目2番3号

④ 代 理 人 弁理士 大岩 増雄 外2名

明 細 書

1. 発明の名称

半導体記憶装置

2. 特許請求の範囲

(1) 行数および列数に配列された複数のメモリセルからなりかつ行数列単位の複数のブロックに分割された第1のメモリセルアレイ、

前記複数のメモリセルの各行を選択するための行選択手段、

前記複数のメモリセルの各列を選択するための列選択手段、

ブロック選択信号を入力するブロック選択信号入力手段、

前記ブロック選択信号に応じて前記第1のメモリセルアレイの前記複数のブロックのいずれかを選択するためのブロック選択手段、

行数および列数に配列された複数のスタティック形メモリセルからなりかつ行数列単位の複数の領域に分割された第2のメモリセルアレイ、

領域選択信号を入力する領域選択信号入力手段、

前記領域選択信号に応じて前記第2のメモリセルアレイの前記複数の領域のいずれかを選択するための領域選択手段、

前記ブロック選択手段により選択された前記第1のメモリセルアレイにおけるブロックと前記領域選択手段により選択された前記第2のメモリセルアレイにおける領域との間でデータの転送を行なうデータ転送手段、

前記第2のメモリセルアレイの前記各領域内において、複数の前記スタティック形メモリセルに対応する情報のうちのいずれかを選択する第1の選択手段、および

前記第1の選択手段により前記各領域ごとに選択された複数の前記情報のいずれかを前記領域選択信号に応じて選択する第2の選択手段を備えた半導体記憶装置。

(2) 前記行選択手段は行アドレス信号に応じて前記第1のメモリセルアレイの行を選択し、前記列選択手段は列アドレス信号に応じて前記第1のメモリセルアレイの列を選択し、

前記第1の選択手段は前記列アドレス信号に応じて前記第2のメモリセルアレイの行および列を選択し、前記第2のメモリセルアレイの各行と前記第1のメモリセルアレイの同一のブロックにおける複数行との間でデータの転送が行なわれる特許請求の範囲第1項記載の半導体記憶装置。

(3) 前記第1のメモリセルアレイは、ダイナミック形メモリセルにより構成され、前記第1のメモリセルアレイからのデータを出力する出力端子と前記第2のメモリセルアレイからのデータを出力する出力端子とを有した特許請求の範囲第1項または第2項記載の半導体記憶装置。

3. 発明の詳細な説明

〔産業上の利用分野〕

この発明は、簡易キャッシュシステム用半導体記憶装置に関し、特にキャッシュメモリを同一チップ上に集積化した半導体記憶装置に関するものである。

〔従来の技術〕

従来より、コンピュータシステムのコストパー

フォーマンスを向上させるために、低速で大容量したがって低コストのDRAMで構成したメインメモリと中央演算処理装置(CPU)との間に、高速のバッファとして小容量の高速メモリを設けることがよく行なわれている。この高速のバッファはキャッシュメモリと呼ばれ、CPUが必要とする可能性の高いデータのブロックが、メインメモリからコピーされて記憶されている。CPUがアクセスしようとしたDRAMのアドレスに記憶されているデータがキャッシュメモリに存在するときにはヒットと呼ばれ、CPUは高速のキャッシュメモリに対してアクセスする。一方、CPUがアクセスしようとしたアドレスに記憶されているデータがキャッシュメモリに存在しないときにはキャッシュミスと呼ばれ、CPUは低速のメインメモリにアクセスすると同時に、そのデータの属するブロックをキャッシュメモリに転送する。

しかしながら、このようなキャッシュメモリシステムは、高価な高速メモリを必要とするので、コストを重視する小型のシステムでは使用すこ

とができなかった。そこで従来は、汎用のDRAMが有しているページモードまたはスタティックコラムモードを利用して簡易キャッシュシステムを構成していた。

第5図はページモードまたはスタティックコラムモードが可能な従来のDRAM素子の基本構成を示すブロック図である。

図において、メモリセルアレイ1には、複数のワード線および複数のビット線対が互いに交差するように配置されており、それらの各交点にメモリセルが設けられている。メモリセルアレイ1のワード線はワードドライバ2を介して行デコード部3に接続されている。またメモリセルアレイ1のビット線対はセンスアンプ部4およびI/Oスイッチ部5を介して列デコード部6に接続されている。行デコード部3には行アドレスバッファ7が接続され、列デコード部6には列アドレスバッファ8が接続されている。これらの行アドレスバッファ7および列アドレスバッファ8には、行アドレス信号RAおよび列アドレス信号CAをマル

チプレクスしたマルチプレクスアドレス信号MPXAが与えられる。さらにI/Oスイッチ部5には出力バッファ9および入力バッファ10が接続されている。

第6A図、第6B図、および第6C図にそれぞれDRAMの通常の読出サイクル、ページモードサイクルおよびスタティックコラムモードサイクルの動作波形図を示す。

第6A図に示す通常の読出サイクルにおいては、まず、行アドレスバッファ7が、行アドレスストローブ信号RASの降下エッジでマルチプレクスアドレス信号MPXAを取込んで行アドレス信号RAとして行デコード部3に与える。行デコード部3はその行アドレス信号RAに応じて、複数のワード線のうち1本を選択する。これにより、この選択されたワード線に接続された複数のメモリセル内の情報が各ビット線に読出され、その情報がセンスアンプ部4により検知、増幅される。この時点で、1行分のメモリセルの情報がセンスアンプ部4にラッチされている。次に、列アドレス

バッファ8が、コラムアドレスストローブ信号 \overline{CAS} の降下エッジでマルチプレクスアドレス信号MPXAを取込んで列アドレス信号CAとして列デコード部6に与える。列デコード部6は、その列アドレス信号CAに応じて、センスアンプ部4にラッチされている1行分の情報のうち1つを選択する。この選択された情報はI/Oスイッチ部5および出力バッファ9を介して出力データD_{OUT}として外部に取出される。この場合のアクセスタイム(RASアクセスタイム) t_{RAC} は、ロウアドレスストローブ信号 \overline{RAS} の降下エッジから出力データD_{OUT}が有効となるまでの時間である。また、この場合のサイクルタイム t_c は、素子がアクティブ状態となっている時間と \overline{RAS} プリチャージ時間 t_{RP} との和となり、概率的な値としては、 $t_{RAC} = 100\text{ns}$ の場合で $t_c = 200\text{ns}$ 程度となっている。

第6B図および第6C図に示すページモードおよびスタティックコラムモードは、同一行上のメモリセルを列アドレス信号CAを変化させてアク

セスするものである。ページモードにおいては、コラムアドレスストローブ信号 \overline{CAS} の降下エッジで列アドレス信号CAをラッチし、スタティックコラムモードにおいては、スタティックRAM(SRAM)のように列アドレス信号CAの変化のみでアクセスする。ページモードおよびスタティックコラムモードの \overline{CAS} アクセスタイム t_{CAC} およびアドレスアクセスタイム t_{AAC} は \overline{RAS} アクセスタイム t_{RAC} のほぼ1/2の値となり、 $t_{RAC} = 100\text{ns}$ に対して 50ns 程度となる。この場合、サイクルタイムも高速になり、ページモードの場合は \overline{CAS} プリチャージ時間 t_{CP} の値によるが、スタティックコラムモードと同様の 50ns 程度の値が得られている。

第7図は、第5図のDRAM素子のページモードあるいはスタティックコラムモードを利用した簡易キャッシュシステムの構成を示すブロック図である。また第8図は第7図の簡易キャッシュシステムの動作波形図である。

第7図において、メインメモリ20は1M×1

構成の8個のDRAM素子21により1Mバイトに構成されている。この場合、行アドレス信号RAと列アドレス信号CAとは合計20ビット(2²⁰ = 1048576 = 1M)必要となる。アドレスマルチプレクサ22は、10ビットの行アドレス信号RAと10ビットの列アドレス信号CAとを2回に分けてメインメモリ20に与えるものであり、20ビットのアドレス信号を受ける20本のアドレス線A₀～A₁₉とマルチプレクスされた10ビットのアドレス信号(マルチプレクスアドレス信号MPXA)をDRAM素子21に与える10本のアドレス線A₀～A₉を有している。アドレスジェネレータ23は、CPU24が必要とするデータに対応するアドレス信号を発生する。ラッチ(TAG)25は、前のサイクルで選択されたデータに対応する行アドレス信号RAを保持しており、コンパレータ26は、20ビットのアドレス信号のうち10ビットの行アドレス信号RAと、TAG25に保持されている行アドレス信号RALとを比較する。両者が一致すれば、

前のサイクルと同じ行がアクセスされた(ヒットした)ことになり、コンパレータ26は高レベルのキャッシュヒット(Cache Hit)信号CHを発生する。ステートマシン27は、キャッシュヒット信号CHに反応して、ロウアドレスストローブ信号 \overline{RAS} を低レベルに保ったままコラムアドレスストローブ信号 \overline{CAS} をトグルするページモード制御を行ない、それに対応してアドレスマルチプレクサ22はDRAM素子21に列アドレス信号CAを与える(第8図参照)。このようにヒットした場合には、DRAM素子21からアクセスタイム t_{CAC} で高速に出力データが得られることになる。

一方、アドレスジェネレータ23から発生された行アドレス信号RAとTAG25が保持していた行アドレス信号RALとが不一致のとき、前のサイクルと異なる行がアクセスされた(キャッシュミスした)ことになり、コンパレータ26は高レベルのキャッシュヒット信号CHを発生しない。この場合、ステートマシン27は通常の読出サイ

クルの \overline{RAS} および \overline{CAS} 制御を行ない、アドレスマルチプレクサ22は行アドレス信号RAおよび列アドレス信号CAを順にDRAM素子21に与える(第8図参照)。このようにキャッシュミスした場合には、 \overline{RAS} のプリチャージから始まる通常の読出サイクルを行ない、低速のアクセスタイム t_{acc} で出力データが得られることになるので、ステートマシン27はウェイト信号Waitを発生し、CPU24に待機をかける。キャッシュミスの場合、TAG25に新しい行アドレス信号RAが保持される。

このように、第7図の簡易キャッシュシステムにおいては、DRAM素子のメモリセルアレイの1行分(1Mビット素子の場合には1024ビット)のデータが1ブロックとなるので、ブロックサイズが不必要に大きく、TAG25に保持されるブロック数(エントリ数)が不足する(第7図のシステムでは1エントリ)ことになり、キャッシュのヒット率が低いという問題点があった。

なお、その他の従来例として、米国特許第4、

577,293号に開示されたような簡易キャッシュシステムもあるが、この簡易キャッシュシステムは1行分のデータを保持するレジスタをメモリセルアレイ外に設け、ヒットした場合は直接このレジスタからデータを取り出すことによりアクセスの高速化を図ったものである。しかしながら、この特許公報に開示された簡易キャッシュシステムも、外部レジスタはメモリセルアレイの1行分のデータを保持するものであり、ブロックサイズが不必要に大きく、第5図および第7図に示す従来例と同様に、キャッシュのヒット率が低いという問題を生ずる。

そこで提案されたのが第9図に示すキャッシュメモリ内蔵DRAM素子である。

このDRAM素子が第5図のDRAM素子と異なるのは以下の点にある。すなわち、DRAMメモリセルアレイ1は、そのアドレス空間上で複数列のメモリセルからなる複数のブロックに分割されている。第9図においては4つのブロックB1~B4に分割されている。そして、センスアンプ

部4とI/Oスイッチ部5との間にトランスファゲート部11およびSRAMメモリセルアレイ12が設けられ、さらにブロックデコード13およびウエイデコード14が設けられている。ブロックデコード13には、ブロック数に応じて列アドレスバッファ8から列アドレス信号CAの一部が供給されるが、その活性化はキャッシュヒット信号CHにより制御される。また、ウエイデコード14には、ウエイアドレスバッファ15を介してウエイアドレス信号WAが与えられる。ウエイデコード14はウエイアドレス信号WAに応じてSRAMメモリセルアレイ12のワード線を選択駆動する。

第10図は第9図のDRAM素子の一部分の構成を詳細に示した図である。

第10図において、センスアンプ部4、トランスファゲート部11、SRAMメモリセルアレイ12、I/Oスイッチ部および列デコード部6は、DRAMメモリセルアレイ1の複数のビット線対B1、 $\overline{B1}$ に対応して、それぞれ複数のセンスア

ンプ40、トランスファゲート110、SRAMメモリセル120、I/Oスイッチ50および列デコード60からなる。また、DRAMメモリセルアレイ1の各ブロックに対応してブロックデコード13が配置されている。各センスアンプ40は各ビット線対B1、 $\overline{B1}$ 間に接続されている。そして各ビット線対B1、 $\overline{B1}$ はNチャネルMOSFETQ1、Q2からなるトランスファゲート110を介してSRAMメモリセルアレイ12のビット線対SBL、 \overline{SBL} に接続されている。SRAMメモリセルアレイ12のビット線対SBL、 \overline{SBL} はNチャネルMOSFETQ3、Q4を介してそれぞれI/OバスI/O、 $\overline{I/O}$ に接続されている。トランスファゲート110のMOSFETQ1、Q2のゲートには、ブロックデコード13により各ブロックごとに共通の転送信号が与えられる。また、各I/Oスイッチ50のMOSFETQ3、Q4のゲートには、対応する列デコード60によりコラム選択信号が与えられる。

このDRAM素子においては、ブロックデコー

ダ13が各ブロックに対応するトランスファゲート110に転送信号を与えることにより、DRAMメモリセルアレイ1からブロック単位で同一行上のデータがSRAMメモリセルアレイ12に転送される。ウエイデコード14によりSRAMメモリセルアレイ12のワード線 $W_1 \sim W_n$ のいずれかが選択されると、そのワード線に接続されたSRAMメモリセル120に記憶されたデータが各ビット線対 $SB_L, \overline{SB_L}$ 上に読出される。ビット線対 $SB_L, \overline{SB_L}$ 上に読出されたデータは、列デコード60から1/0スイッチ50にコラム選択信号が与えられることによって、1/0バス $1/0, \overline{1/0}$ に読出される。

このDRAM素子によると、複数列の1行のデータを1つのデータブロックとして、異なる行上の複数のデータブロックが複数のSRAMメモリセル120に保持される上に、同一列の異なる行上のデータブロックが同時にSRAMメモリセルアレイ12上に保持される(アソシアティビティ)。したがって、このSRAMメモリセルアレイを

6の数が増加している点、および、コンパレータ26からの出力であるキャッシュヒット信号CHおよびウエイアドレス信号WAがDRAM素子31に入力されている点である。ここでは、ウエイアドレス信号は2ビットである。

第11図の簡易キャッシュシステムの動作を従来の簡易キャッシュシステムの説明で用いた第6A図～第6C図および第12図の動作波形図を参照しながら説明する。

TAG25には、各ブロック別に最も新しいサイクルで選択された行に対応する行アドレスが複数組キャッシュ用アドレスセットとして保持されている。ここでは、ウエイアドレス信号として2ビットを考えているので、4組の行アドレスが保持されている。したがって、ブロック数を4とすると16組のアドレスセットがTAG25に記憶されていることになる。また、よく使用されるアドレスの組を固定的にTAG25に保持させておいてもよい。

まず、CPU24が必要とするデータに対応す

キャッシュメモリとして利用すれば、データのエンタリ致を増すことができ、その結果、キャッシュのヒット率を向上することができる。

さらに、SRAMメモリセルアレイ12のワード線 $W_1 \sim W_n$ を非活性状態に保っておけば、DRAMメモリセルアレイ1への寄込動作時やDRAMメモリセルアレイ1からの読出動作時にも、キャッシュメモリへの転送を行なわない構成が可能となり、キャッシュメモリシステムへの応用に自由度が増すという利点が生じる。

第11図は第9図のDRAM素子を利用した簡易キャッシュシステムの構成を示すブロック図である。

第11図において、メインメモリ30は1M×1構成の8組のDRAM素子31により1Mバイトに構成されている。第11図のメモリシステムが第7図のメモリシステムと相違するのは、DRAM素子31のブロック分けの数およびSRAMメモリセルアレイ12のワード線の本数(セット数)に対応してTAG25およびコンパレータ2

るアドレス信号をアドレスジェネレータ23が発生する。コンパレータ26は、20ビットのアドレス信号のうち10ビットの行アドレス信号RAおよび列アドレス信号CAのうちブロック分けに相当する複数ビット(第9図に示す例では2ビット)と、TAG25に保持されたアドレスセットとを比較する。そして両者が一致すればキャッシュにヒットしたことになり、コンパレータ26は高レベルのキャッシュヒット信号CHおよびヒットしたブロックのウエイアドレス信号WAを発生する。ステータマシン27は、このキャッシュヒット信号CHに反応して、ロウアドレスストローブ信号RASを低レベルに保ったままコラムアドレスストローブ信号CASをトグルし、これに反応してアドレスマルチプレクサ22はDRAM素子31に10ビットの列アドレス信号CAを与える(第12図参照)。このとき、DRAM素子31においては、第9図に示したようにキャッシュヒット信号CHによる制御により、列アドレス信号CAはブロックデコード13には供給されない。

したがって、DRAMメモリセルアレイ1とSRAMメモリセルアレイ12とは分離された状態を保つ。そして、ウエイアドレス信号WAに対応した1行分のSRAMメモリセル120から各ビット線対SBL, $\overline{\text{SBL}}$ 上にデータが読出される。また、列アドレス信号CAに応じた1/0スイッチ50が、列デコード60によって導通状態にされる。これにより、列アドレス信号CAおよびウエイアドレス信号WAに対応するSRAMメモリセル120内のデータが1/0バス1/0, $\overline{\text{T}}\overline{\text{A}}$ および出力バッファ9を介して出力される。このようにヒットした場合には、SRAMメモリセル120からページモードのようにアクセスタイムt_{acc}で高速に出力データが得られることになる。

一方、アドレスジェネレータ23から発生されたアドレス信号とTAG25に保持されたキャッシュ用アドレスセットとが不一致のときは、キャッシュミスしたことになり、コンパレータ26は高レベルのキャッシュヒット信号CHを発生しな

い。この場合、ステートマシン27は通常の読出サイクルのRASおよびCAS制御を行ない、アドレスマルチプレクサ22は行アドレス信号RAおよび列アドレス信号CAを順にDRAM素子31に供給する(第12図参照)。このようにキャッシュミスした場合には、低速のアクセスタイムt_{acc}で出力データが得られることになるので、ステートマシン27はウエイ信号Waitを発生し、CPU24に待機をかける。キャッシュミスの場合は、そのときにアクセスされたメモリセルを含むブロックのデータが、ブロックデコード13により導通状態とされるトランスファゲート110を介して、DRAMメモリセルアレイ1のビット線BL, $\overline{\text{BL}}$ から、ウエイアドレス信号WAにより選択されたSRAMメモリセル120のブロックに一括転送される。これにより、このブロックのSRAMメモリセル120の記憶内容が替換えられる。また、そのブロックの対応するウエイアドレス信号WAに関するTAG25には新しいアドレスセットが保持される。

このように、第9図のDRAM素子を用いた簡易キャッシュシステムにおいては、キャッシュメモリとしてのSRAMメモリセルアレイ12に複数のブロックのデータが保持されるので、TAG25へのデータのエントリ数を増加することが可能となり、キャッシュのヒット率が高くなる。

また、ここでは、キャッシュミスした場合に、DRAMメモリセルアレイにアクセスすると同時に、SRAMメモリセルアレイからなるキャッシュメモリにデータを転送する例を示したが、SRAMメモリセルアレイのすべてのワード線を非選択状態にすることでこの転送を禁止することもできる。同様に、DRAMメモリセルアレイへの送込動作の場合も、SRAMメモリセルアレイへ転送するか否かを選択することも可能である。なお、第11図に示した例は、4ウエイセットアソシティブキャッシュシステムに相当する。

【発明が解決しようとする問題点】

しかし、上記の簡易キャッシュシステムにおいては、キャッシュヒットした場合、キャッシュメ

モリとしてのSRAMメモリセルアレイ12をアクセスするためのアドレス信号のうちウエイアドレス信号WAは、コンパレータ26での比較後に出力される。したがって、ウエイアドレス信号WAのDRAM素子31への供給が遅れるため、SRAMメモリセルアレイ12のワード線の駆動が遅れ、高速のSRAMメモリセルアレイ12をキャッシュメモリとして使用できる装置でありながら、ヒット時のアクセスタイムを高速にできないという欠点があった。

この発明は上記のような問題点を解消するためになされたもので、ヒット時のアクセスタイムを高速にできるキャッシュシステムを構成することが可能なキャッシュメモリ内蔵半導体記憶装置を提供することを目的としている。

【問題点を解決するための手段】

この発明に係る半導体記憶装置は、第1のメモリセルアレイ、行選択手段、列選択手段、ブロック選択信号入力手段、ブロック選択手段、第2のメモリセルアレイ、領域選択信号入力手段、領域

選択手段、データ転送手段、第1の選択手段、および第2の選択手段を備えたものである。

第1のメモリセルアレイは、複数行および複数列に配列された複数のメモリセルからなり、かつ複数列単位の複数のブロックに分割されている。行選択手段は、複数のメモリセルの各行を選択するためのものである。列選択手段は、複数のメモリセルの各列を選択するためのものである。ブロック選択信号入力手段は、ブロック選択信号を入力するためのものである。ブロック選択手段は、ブロック選択信号にตอบสนองして、第1のメモリセルアレイの複数のブロックのいずれかを選択するものである。

また、第2のメモリセルアレイは、複数行および複数列に配列された複数のスタティック形メモリセルからなり、かつ複数列単位の複数の領域に分割されている。領域選択信号入力手段は、領域選択信号を入力するためのものである。領域選択手段は、領域選択信号にตอบสนองして、第2のメモリセルアレイの複数の領域のいずれかを選択するも

のである。データ転送手段は、ブロック選択手段により選択された第1のメモリセルアレイにおけるブロックと領域選択手段により選択された第2のメモリセルアレイにおける領域との間でデータの転送を行なうものである。

さらに、第1の選択手段は、第2のメモリセルアレイの各領域内において複数のスタティック形メモリセルに対応する情報のうちのいずれかを選択するものである。第2の選択手段は、第1の選択手段により各領域ごとに選択された複数の情報のいずれかを領域選択信号にตอบสนองして選択するものである。

【作用】

この発明に係るキャッシュメモリ内蔵半導体記憶装置においては、第1のメモリセルの複数行上の複数列単位のデータブロックを第2のメモリセルアレイ上に保持することができる上に、第1のメモリセルアレイの同一列における異なる行のデータブロックを複数列同時に第2のメモリセルアレイの異なる領域に保持することができる。また、

第1のメモリセルアレイの同一列にかかる異なる行のデータブロックを第2のメモリセルアレイの同一行に配置することもできる。したがって、この第2のメモリセルアレイをキャッシュメモリとして利用すれば、データのエントリ数を効率的に増すことができ、キャッシュのヒット率を向上させることができるばかりでなく、キャッシュメモリのアクセスタイムを高速にすることができる。

【実施例】

以下、この発明の一実施例を図面を用いて説明する。

第1図はこの発明の一実施例によるDRAM素子の構成を示すブロック図である。

この実施例は以下の点を除いて第9図に示すDRAM素子と同様であり、相当部分には同一の参照番号を付し、適宜その説明を省略する。

図において、DRAMメモリセルアレイ1は、そのアドレス空間上で複数のブロックに分割されている。この実施例では4つのブロックBK1～BK4に分割されている。一方、SRAMメモリ

セルアレイ12は複数列単位の複数のウエイに分割されている。この実施例では4つのウエイA～Dに分割されている。但し、DRAMメモリセルアレイ1のブロック数とSRAMメモリセルアレイ12のウエイ数は異なってもよい。

DRAMメモリセルアレイ1とSRAMメモリセルアレイ12との間には、センスアンプ部4、ブロックトランスファゲート部11、内部1/0部41、およびウエイトランスファゲート部42が配置されている。ブロックトランスファゲート部11は、DRAMメモリセルアレイ1のいずれかのブロックの1行のデータを内部1/0部41に転送するものである。ブロックデコード13は、列アドレス信号CAのうちの一部（この実施例の場合2ビット）にตอบสนองして、DRAMメモリセルアレイ1のどのブロックのデータを転送するかをブロックトランスファゲート部11に指令するものである。ウエイトランスファゲート部42は、内部1/0部41に転送されたデータを、SRAMメモリセルアレイ12のいずれかのウエイに転

送するものである。ウエイデコード14は、ウエイアドレスバッファ15を介して与えられるウエイアドレス信号WAに应答して、内部I/O帯41のデータをSRAMメモリセルアレイ12のどのウエイに転送するかをウエイトランスファゲート部42に指令するものである。

SRAMメモリセルアレイ12には、キャッシュ行デコード43、キャッシュI/Oスイッチ部44およびキャッシュ列デコード部45が設けられている。キャッシュ行デコード43は、キャッシュアドレスバッファ46から与えられるキャッシュ行アドレス信号に应答して、SRAMメモリセルアレイ12の1行を選択するものである。キャッシュ列デコード部45は、キャッシュアドレスバッファ46から与えられるキャッシュ列アドレス信号に应答して、各ウエイ内の1列を選択するものである。キャッシュアドレスバッファ46は、DRAMメモリセルアレイ1に与えられる列アドレス信号CAをキャッシュアドレス信号CCAとして入力し、その一部をキャッシュ行デコー

ダ43にキャッシュ行アドレス信号として与え、他をキャッシュ列デコード45にキャッシュ列アドレス信号として与えるものである。キャッシュI/Oスイッチ部44には、SRAMメモリセルアレイ12の各ウエイに対応する複数のSRAM用センスアンプ47がそれぞれI/O線対I/O₁〜I/O_nを介して接続されている。

キャッシュ行デコード43およびキャッシュ列デコード部45により各ウエイごとに選択されたSRAMメモリセルアレイ12内のデータがそれぞれ対応するSRAM用センスアンプ47により検知、増幅される。ウエイセクタ48は、ウエイアドレスバッファ15から与えられるウエイアドレス信号WAに应答して、複数のSRAM用センスアンプ47により与えられたデータのうちの1つを選択して、出力バッファ9bを介してキャッシュ出力データD_{out}として外部に出力するものである。キャッシュ入力データD_{in}として入力バッファ10bに与えられたデータをSRAMメモリセルアレイ12の1つのメモリセルに書

込む場合は、上記と逆の経路で行なわれる。

第1図においては、DRAMメモリセルアレイ1のブロックBK1の各行のデータA₁、B₁、C₁およびD₁がSRAMメモリセルアレイ12の各ウエイA、B、CおよびDの同一行にそれぞれ転送された状態が示されている。

第2図は、第1図の一部分の構成を詳細に示す図である。

DRAMメモリセルアレイ1の各ブロックBK1〜BK4において、センスアンプ部4およびブロックトランスファゲート部11は、n組のビット線対BL₁〜BL_nに対応してそれぞれn個のセンスアンプ部40およびn個のブロックトランスファゲート110からなる。また、内部I/O帯41は、n組のI/O線対I/O₁〜I/O_nからなる。各ブロックのビット線対BL₁〜BL_nはセンスアンプ40およびブロックトランスファゲート110を介して対応するI/O線対I/O₁〜I/O_nにそれぞれ接続されている。

一方、SRAMメモリセルアレイ12は4つの

ウエイに分割され、各ウエイはn列のSRAMメモリセル120、すなわちn組のビット線対SBL₁〜SBL_nからなる。各ウエイにおいて、ウエイトランスファゲート部42は、n組のビット線対SBL₁〜SBL_nに対応してそれぞれn個のウエイトランスファゲート420からなる。各ウエイにおけるn組のビット線対SBL₁〜SBL_nは、それぞれウエイトランスファゲート420を介して内部I/O帯41の対応するI/O線対I/O₁〜I/O_nにそれぞれ接続されている。キャッシュI/Oスイッチ部44は、SRAMメモリセルアレイ12の各ビット線対SBL₁〜SBL_nに対応する複数のキャッシュI/Oスイッチ440および各ウエイに対応する4組のI/O線対I/O₁〜I/O_nからなる。各ウエイに属するn組のビット線対SBL₁〜SBL_nは、それぞれキャッシュI/Oスイッチ440を介して、そのウエイに対応するI/O線に接続されている。たとえば、ウエイCに属するビット線対SBL₁〜SBL_nはすべてI/O線対I/O₁〜I/O_nに接続さ

れている。また、各ウエイごとにキャッシュ列デコード部45が設けられている。各ウエイのキャッシュ列デコード部45は、各列に対応するn個のキャッシュ列デコード450からなる。各キャッシュ列デコード450は、対応するキャッシュ1/Oスイッチ440のMOSトランジスタのゲートに接続されている。

第3図は、第1図のDRAM素子を利用した簡易キャッシュシステムの構成を示すブロック図である。

第3図において、メインメモリ30は1M×1構成の8個のDRAM素子31により1Mバイトに構成されている。第3図のメモリシステムが第11図のメモリシステムと相違するのは、コンパレータ26からの出力であるキャッシュヒット信号CHの代わりに、マルチプレクサ22によりマルチプレクスされる前の列アドレス信号に相当する10ビットのアドレス信号がキャッシュアドレス信号CCAとしてDRAM素子31に入力されている点、および、キャッシュヒット信号CHに

応答してステートマシン27が発生するデータセレクト信号DSがデータセクタ51に入力されている点である。データセクタ51は、データセレクト信号DSに反応して、DRAM素子31から与えられるDRAMデータDDまたはキャッシュデータCDを選択して出力するものである。

第3図の簡易キャッシュシステムの動作を第4図に示す動作波形図を参照しながら説明する。

TAG25には、各ブロック別に最も新しいサイクルで選択された行に対応する行アドレスが複数組キャッシュ用アドレスセットとして保持されている。ここでは、ウエイアドレス信号WAとして2ビットを考えているので、4組の行アドレスが保持されている。したがって、ブロックを4とすると16組のアドレスセットがTAG25に記憶されていることになる。また、よく使用されるアドレスを固定的にTAG25に保持させておいてもよい。

まず、CPU24が必要とするデータに対応するアドレス信号をアドレスジェネレータ23が発

生する。コンパレータ26は、20ビットのアドレス信号のうち10ビットの行アドレス信号RAおよび列アドレス信号CAのうちブロック分けに相当する複数ビット（第3図に示す例では2ビット）とTAG25に保持されたアドレスセットとを比較する。そして両者が一致すればキャッシュがヒットしたことになり、コンパレータ26は高レベルのキャッシュヒット信号CHおよびヒットしたブロックのウエイアドレス信号WAを発生する。

このコンパレータ26によるアドレス信号の比較に先立って、キャッシュヒットすることを前提に、DRAM素子31へは10ビットのキャッシュアドレス信号CCAが入力され、SRAMメモリセルの読出動作が進行している。ここでは4ウエイを考えているので4ビットの読出動作が進行している。したがって、キャッシュにヒットしたときは、ウエイアドレス信号WAが入力されると、高速に所望のデータがキャッシュデータCDとしてキャッシュ出力バッファ9bを介して出力され、

キャッシュヒット信号CHに反応して発生されるデータセレクト信号DSによって、データセクタ51からキャッシュメモリのデータが得られることになる。

逆に、コンパレータ26に入力されたアドレス信号がTAG25に保持されたアドレスセットと不一致のときは、キャッシュミスしたことになり、コンパレータ26はキャッシュヒット信号CHを発生しない。これにより、SRAMメモリセルから出力されるキャッシュデータCDは無視されることになる。この場合、ステートマシン27は通常の読出サイクルの \overline{RAS} および \overline{CAS} 制御を行ない、アドレスマルチプレクサ22は行アドレス信号RAおよび列アドレス信号CAを順にDRAM素子31に供給する（第4図参照）。このようにキャッシュミスした場合には、低速のアクセスタイムt_{acc}で出力データが得られることになるので、ステートマシン27はウエイ信号Weiiを発生し、CPU24に待機をかける。キャッシュミスの場合は、そのときにアクセスされた

メモリセルを含むブロックのデータが、ブロックデコード13により導通状態とされるブロックトランスファゲート110を介して、内部1/0帯41の1/0線対1/0、 \sim 1/0に転送される。そして、それらのデータはウエイアドレス信号WAにより選択されるウエイトランスファゲート420を介してSRAMメモリセルアレイ12の適当なウエイに転送され、キャッシュ行デコード43により選択された行上のSRAMメモリセル120の記憶内容が替换えられる。また、そのデータのブロックの対応するウエイに関するTAG25には、今回アクセスされた新しいアドレスセットが保持される。

以上説明したように、上記実施例では、キャッシュメモリとしてのSRAMメモリセルアレイ120に複数ブロック分のデータが保持されるので、TAG25へのデータのエントリ数を増すことができ、その結果ヒットの確率を向上させることができ、かつ、キャッシュメモリのアクセスタイムが高速になるという効果がある。

装置の構成を示すブロック図、第2図は第1図の半導体記憶装置の一部分の構成を詳細に示すブロック図、第3図は第1図の半導体記憶装置を利用した簡易セットアソシアティブキャッシュシステムの構成を示すブロック図、第4図は第3図の簡易キャッシュシステムの動作波形図、第5図は従来のDRAM素子の構成を示すブロック図、第6A図は従来のDRAM素子の通常の読出サイクルの動作波形図、第6B図は従来のDRAM素子のページモードサイクルの動作波形図、第6C図は従来のDRAM素子のスタティックコラムモードサイクルの動作波形図、第7図は第5図のDRAM素子を利用した簡易キャッシュシステムの構成を示すブロック図、第8図は第7図の簡易キャッシュシステムの動作波形図、第9図はキャッシュメモリ内蔵DRAM素子の構成を示すブロック図、第10図は第9図のDRAM素子の一部分の構成を詳細に示すブロック図、第11図は第9図のDRAM素子を利用した簡易キャッシュシステムの構成を示すブロック図、第12図は第11図の簡

【発明の効果】

以上のようにこの発明によれば、ブロックサイズを不必要に大きくすることなく、第1のメモリセルアレイのデータブロックを第2のメモリセルアレイに多数保持することができるので、データのエントリ数を効率的に増加することができる。さらに、第1のメモリセルアレイの同一列に関する異なる行のデータブロックを第2のメモリセルアレイ上で同一行に存在するようにし、第2のメモリセルアレイの各領域から情報を読出した後に領域のいずれかを選択することができるので、第2のメモリセルアレイをキャッシュメモリとして用いると、キャッシュヒットした場合のアクセスタイムを飛躍的に高速化することが可能となる。したがって、この発明の半導体記憶装置を用いればキャッシュのヒット率が高く高速な簡易セットアソシアティブキャッシュシステムを構成することができる。

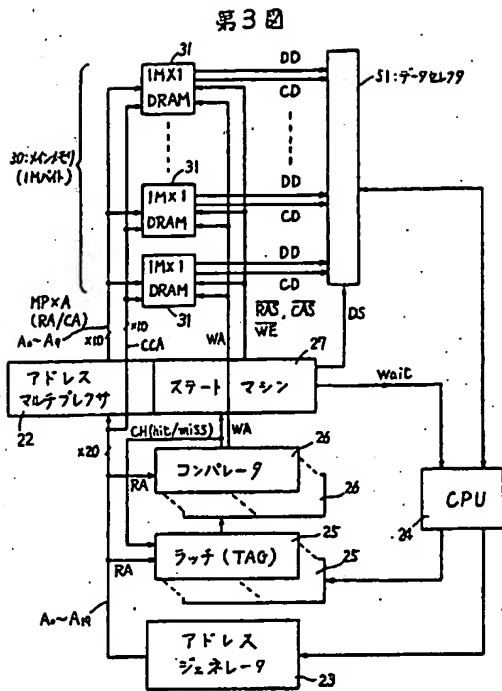
4. 図面の簡単な説明

第1図はこの発明の一実施例による半導体記憶

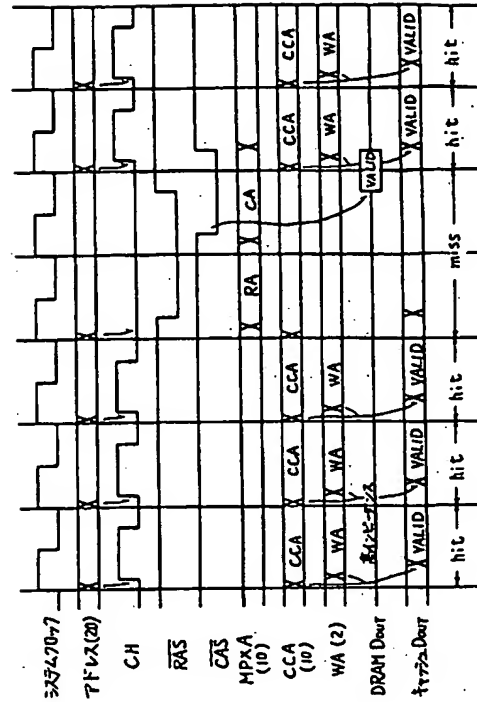
品キャッシュシステムの動作波形図である。

図において、1はDRAMメモリセルアレイ、2はワードドライバ、3は行デコード部、4はセンスアンプ部、5は1/0スイッチ部、6は列デコード部、7は行アドレスバッファ、8は列アドレスバッファ、9a、9bは出力バッファ、10a、10bは入力バッファ、11はブロックトランスファゲート部、12はSRAMメモリセルアレイ、13はブロックデコード、14はウエイデコード、15はウエイアドレスバッファ、41は内部1/0帯、42はウエイトランスファゲート部、43はキャッシュ行デコード、44はキャッシュ1/0スイッチ部、45はキャッシュ列デコード部、46はキャッシュアドレスバッファ、47はSRAM用センスアンプ、48はウエイセレクト、BL、 \overline{BL} はDRAMメモリセルアレイのビット線対、SBL、 \overline{SBL} はSRAMメモリセルアレイのビット線対である。

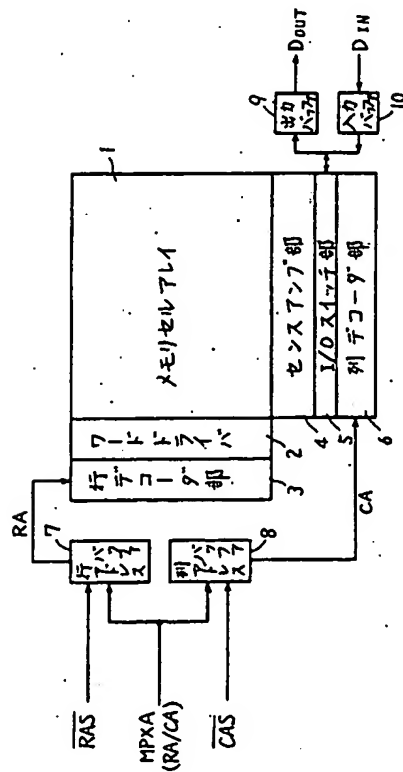
なお、図中、同一番号は同一または相当部分を示す。



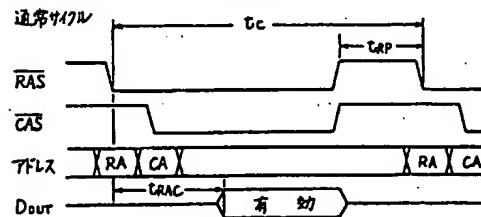
第4図



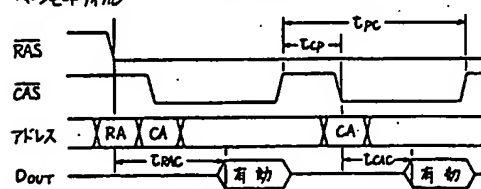
第5図



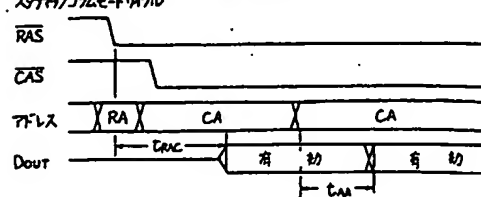
第6A図



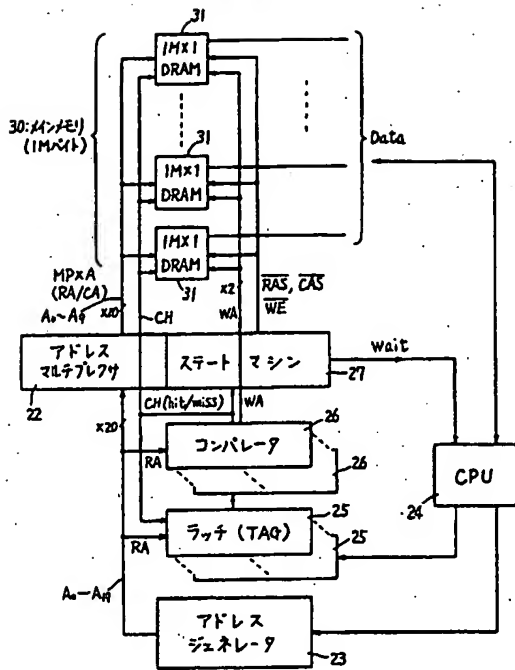
第6B図



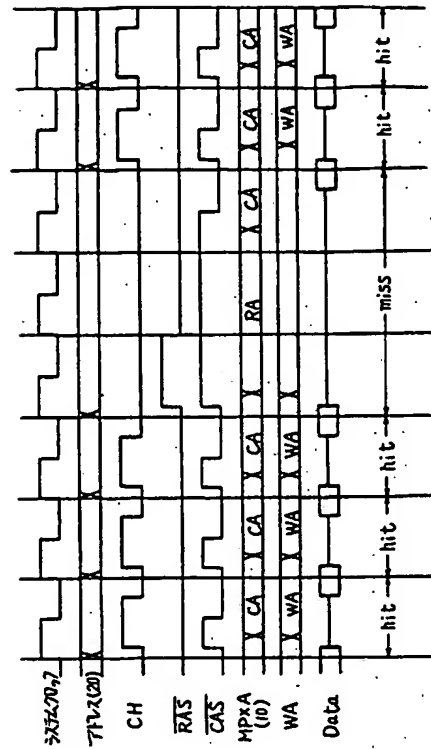
第6C図



第11図



第12図



(19) JAPANESE PATENT OFFICE (JP)

(12) PUBLICATION OF UNEXAMINED (KOKAI) PATENT APPLICATION (A)

(11) Kokai (Laid-Open) Patent Application Number 1-124193

(43) Date of Disclosure: May 17, 1989

(51) Int. Cl. ⁴	Identification Symbol	Intra-Agency Number
G 11 C 11/34	362	C-8522-5B

Examination requested: not yet requested
Number of inventions: 1 (total of 14 pages)

(54) Title of the Invention: SEMICONDUCTOR MEMORY DEVICE

(21) Application Number: 62-281619
(22) Filing Date: November 6, 1987

(72) Inventor: Kazuyasu FUJISHIMA
c/o Mitsubishi Electric Corporation
LSI Research Institute
Hyogo-ken, Itami-shi, Mizuhara 4-chome, 1-banchi

(72) Inventor: Yoshio MATSUDA
c/o Mitsubishi Electric Corporation
LSI Research Institute
Hyogo-ken, Itami-shi, Mizuhara 4-chome, 1-banchi

(72) Inventor: Mikio ASAKURA
c/o Mitsubishi Electric Corporation
LSI Research Institute
Hyogo-ken, Itami-shi, Mizuhara 4-chome, 1-banchi

(71) Applicant: Mitsubishi Electric Corporation
Tokyo-to, Chiyoda-ku, Marunouchi 2-chome, 2-ban, 3-go

(74) Representative: Masuo OIWA, patent attorney, 2 others

Specifications

1. Title of the Invention: Semiconductor Memory Device
2. Scope of the Patent=s Claims

(1) A semiconductor manufacturing device, characterized by the fact that it is equipped with a plurality of memory cells arranged in a plurality of rows and a plurality of columns, having a first memory cell array, which is divided into a plurality of blocks comprising a plurality of column units;

a row selection means, which is used to select respective rows of said plurality of memory cells;

a column selection means, which is used to select respective columns of said plurality of memory cells;

a block selection means, which inputs a block selection signal;

a block selection means, which is used to select any block from said plurality of blocks in said first memory cell array, in response to said block selection signal;

a second memory cell array, comprising a plurality of static type memory cells, arranged in a plurality of rows and a plurality of columns and divided into a plurality of regions in a plurality of column units;

a region selection means, which is used to select one out of said plurality of regions in said second memory cell array in response to said region selection signal;

a data transmission means, transmitting data between the blocks in said first memory cell array, selected with said block selection means, and said second memory cell array, selected with said region selection means;

a first selection means, which can select any information from the information items corresponding to a plurality of said static type memory cells, in each of said regions of said second memory cell array; and

and a second selection means, which can select in response to said region selection signal any item among said plurality of information items selected per each of said regions by said first selection means.

(2) The semiconductor manufacturing device of claim 1, wherein said row selection means selects a row in said first memory cell array in response to the row address signal;

said column selection means selects a column in response to the column address signal;

[page 2]

and said first selection means selects said second memory cell array rows and columns in

response to said column address signal; so that data transmission operations are applied to a plurality of rows in the same block of said first memory cell array and each row of said second memory cell array.

(3) The manufacturing device of claim 1 or claim 2, wherein the construction of said first memory cell array comprises memory cells of the static type, equipped with output terminals providing output of data from said first memory cell array, and with output terminals providing output of data from said second memory cell array.

3. Detailed Explanation of the Invention

(Sphere of Industrial Use)

This invention relates to a semiconductor memory device for a simple cache system, in particular to a semiconductor memory device having a cache memory integrated on the same chip.

(Prior Art Technology)

According to prior art, in order improve the cost performance of a computer system, a high speed memory having a low capacity was deployed as a high-speed buffer between the central processing unit (CPU) and the main memory construction designed with a DRAM construction for low speed and thus with at a low cost. This high-speed buffer is a so called cache memory in which a block of data, which is highly like to be required by the CPU, is copied from the main memory and stored. The state when data which stored in the DRAM address is present in the cache memory so that it could be accesses by the CPU is referred to as Ahit@, as the CPU can access the cache memory with a high speed. On the other hand, the state when the data stored in the address to be accessed by the CPU is not present in the cache memory is referred to as Acache miss@, wherein the CPU accesses the main memory with a low speed, while the CPU at the same time transfers to the cache memory the data block to which this data belongs.

However, because expensive high-speed memory is required for such a cache memory system, the system could not be used in a compact system in view of the cost. That is why conventionally, a general-purpose DRAM was configured as a page mode or static column mode configuration.

Figure 5 is a block diagram showing a basic construction of a conventional DRAM element which can be used with the page mode or with the static column mode.

In this figure, a memory cell array 1 is provided with a plurality of word lines and a plurality of bit lines arranged in a mutually intersecting arrangement, while memory cells are deployed in the intersection points. The word lines of the memory array 1 are connected through a word driver 2 to a row decoder part 3. In addition, the bit lines of the memory cell array 1 are connected through a sense amplifier part 4 and an I/O switch part 5 to a column decoder part 6.

A row address buffer 7 is connected to the row decoder part 3, and a column address buffer 8 is connected to the column decoder part 6. A multiple signal MPXA is applied multiplexed with the row address signal RA and the column address signal CA to the row address buffer 7 and the column address buffer 8. Furthermore, an output buffer 9 and an input buffer 10 are connected to the I/O switch part 5.

Figure 6A, Figure 6B and Figure 6C are waveform diagrams showing the operation of the cycle in an ordinary read cycle, page mode cycle and static column mode cycle of the DRAM, respectively.

In the ordinary read cycle shown in Figure 6A, the row address buffer 7 first acquires the multiplex address signal MPXA at the falling edge of the row address strobe signal RAS and applies it as row address signal RA to the row decoder part 3. The row decoder part 3 selects one word line from a plurality of word lines according to this row address signal RA. Because of that, information stored in a plurality of memory cells connected to this selected word line is read to each word line, and this information is detected and amplified by the sense amplifier part 4. At this point in time, information stored in the memory cell of 1 row segment is latched by the sense amplifier part 4.

[page 3]

Next, the column address buffer 8 acquires the multiplex address signal MPXA at the falling edge of the column address strobe signal CAS and applies it as column address signal CA to the column decoder part 6. The column decoder part 6 then selects one of the information items in 1 line segment latched by the sense amplifier part 4 in response to this column address signal CA. This selected information is extracted to an external part as output data D_{OUT} through the I/O switch part 5 and the output buffer 9. The access time (RAS access time) t_{RAC} is in this case the time period valid from the falling edge of the row address strobe signal RAS until the output data D_{OUT} becomes valid. The cycle time t_c is in this case the sum of the time period created by the active state of the element and of the time period t_{RP} of the RAS precharge. As a standard value, t_c is approximately 200 ns when $t_{RAC} = 100$ ns.

In the page mode and static column mode shown in Figure 6B and Figure 6C, memory cells on the same row are accessed by changing the column address signal CA. In the page mode, the column address signal CA is latched at the falling edge of the column address row signal CAS. In the static column mode, access is achieved by simply changing the column address signal CA as in a static RAM (SRAM). The page mode and static column mode CAS access time CAS and address access time T_{AA} therefore create about one 2 of the value of the RAS access time t_{RAC} for $t_{RAC} = 100$ ns., which is about 50 ns. A high speed of the of the cycle time is also created in this case, and in the case of the page mode, the value of about 50 ns is obtained in the same manner as in the static column mode with the CAS precharge time period value t_{CP} .

Figure 7 is a block diagram showing the construction of a simple cache system using the page mode or static column mode of the DRAM element shown in Figure 5. In addition, Figure 8 is a waveform diagram explaining the operation of the simple cache system shown in Figure 7.

As shown in Figure 7, the construction of the main memory 20 comprises 1 M byte comprising 8 DRAM elements 21. In this case, a total of 20 bits ($2^{20} = 1048576 = 1 \text{ M}$) is required by the row address signal RA and the column address signal CA. An address multiplexer 22, which applies 10-bit row address signal RA and 10-bit column address CA to the memory 20 twice, has 19 address lines $A_0 \sim A_{18}$, applying to the DRAM elements 21 10-bit address signal (multiplex address signal MPXA), which is received as 10-bit address signal and multiplexed by 20 address lines $A_0 \sim A_{19}$.

An address generator 23 generates address signal corresponding to the data requested by the CPU 24. A latch (TAG) 25 holds the row address signal RA corresponding to the data selected in the previous cycle. A comparator 26 compares row address signal RAL held in the TAG 25 to the 10-bit row address signal out of the 20-bit address signal. When both coincide, this means that the same row is accessed (hit) as in the previous cycle, the comparator 26 generates high-level cache hit (Cache Hit) signal CH. A state machine 27, responding to the cache hit signal CH, performs page mode control by toggling the address strobe signal CAS and maintaining a low level of the row address signal RAS. In response to that, the address multiplexer 22 applies the column address signal CA to the DRAM element 21 (see Figure 8). Therefore, in the case of such a $A_{hit@}$, output data will be obtained at a high speed from the DRAM element 21 with the access time t_{CAC} .

On the other hand, when the address signal RA obtained from the address generator 23 does not match the row address signal RAL held by the TAG 25, a different row will be accessed than the row accessed in the previous cycle (cache miss), and the cache hit signal CH will not be generated at the high level by the comparator 26.

[page 4]

In this case, the state machine 27 performs ordinary reading cycle RAS and CAS control and the address multiplexer 22 applies sequentially the row address signal RA and the column address signal CA to the DRAM element 21 (see Figure 8). In the case of such a cache miss, an ordinary read cycle starts from the RAS precharge, and because output data is obtained with low-speed access time t_{RAC} , the state machine 27 generates the wait signal Wait and CPU 24 is brought into a wait state. In the case of a cache miss, a new row address signal RA is held in the TAG 25.

Therefore, because in the simple cache system shown in Figure 7, data corresponding to 1 row segment of the memory cell array of the DRAM element (1024 bits for a 1 M bit element) creates 1 block, the size of the block is unnecessarily large and because the number of blocks

(number of entries) held in TAG 25 is insufficient (1 entry in the system shown Figure 7), this created a problem known as a low cache hit rate.

Moreover, another conventional example of a simple cache system has been disclosed also in US Patent Number 4,577,293. This simple cache system has a register holding data in 1-row segments, which is created outside of the memory cell array, so that in the case of a hit, access is achieved with a high-speed design by reading data directly from this register. However, also according to this simple cache system disclosed in the US Patent Gazette, the block size is unnecessarily large and the problem is that a low cache hit rate is created in the same manner as in the conventional examples shown in Figure 5 and Figure 7.

That is why a DRAM element provided with a built-in cache memory has been proposed as shown in Figure 9.

The differences between this DRAM element and the DRAM element shown in Figure 5 are as follows. Specifically, a DRAM memory array 1 is divided into a plurality of blocks comprising a plurality of memory cells in the address space of the memory array. In Figure 9, the array is divided into 4 blocks B1 ~ B4. Also, a transfer gate part 11 and a SRAM memory cell array 12 are deployed between a sense amplifier part 4 and an I/O switch part 5. Moreover, a block decoder 13 and a window decoder 14 are also used. While one part of the address signal CA is supplied from the column address buffer 8 according to the block number to the block decoder 13, the activation of the operation is controlled by the cache hit signal CH. Also, a way address signal WA is applied through a way buffer 15 to the way decoder 14. The way decoder 14 is operated to select a word line from the SRAM memory cell array 12 according to the way address signal WA.

Figure 10 is a diagram showing the detailed construction in one part of the DRAM element shown in Figure 9.

In Figure 10 is shown a sense amplifier part 4, a transfer gate part 11, a SRAM memory cell array 12, and an I/O switch part and column decoder part 6. The construction further comprises multiple bit line pairs BL, BL of the DRAM memory cell array 1, and the corresponding multiple sense amplifiers 40, transfer gates 110, SRAM memory cells 120, and I/O switches 50 and column decoders 60. In addition, block decoders 13 are arranged to correspond to each block of the DRAM memory cell array 1. Each sense amplifier 40 is connected to each bit line pair BL, BL. Also, each bit line part BL, BL is connected through a pair of bit lines SBL, SBL to the SRAM memory cell array 12 through the transfer gate 110 from the N-channel MOSFETs Q1, Q2. The bit line pairs SBL, SBL of the SRAM memory array 12 are connected to each I/O band I/O, I/O through the N-channel MOSFETs Q3, Q4. A joint transmission signal is applied per each block from the block decoder 13 to the gate of the MOSFETs Q1, Q2 of the transfer gate 110. Also, a column selection signal is applied with a corresponding column decoder 60 to each gate of the MOSFETs Q3, Q4 of each I/O switch 50.

When the transmission signal is applied to the transfer gate corresponding to each block by the block decoder 13 in this DRAM element, data on the same row is transmitted in block units from the DRAM memory array 1 to the SRAM memory cell array 12. When one of the word lines $W_1 \sim W_n$ is selected from the SRAM memory cell array 12 by the way decoder 14, the data in each bit line pair SBL, SBL that is stored in the SRAM memory cell 120 connected to the word lines is read. The data read on the bit line pairs SBL, SBL is read in the I/O band I/O, I/O by applying the column selection signal to the I/O switch 50 from the column decoder 60.

According to this DRAM element, because 1 data block has data in which 1 line corresponds to a plurality of columns, multiple data blocks on different rows are held in multiple SRAM memory cells 120, and the data blocks on different rows in the same column are held in the SRAM memory cell array 12 (associativity).

Accordingly, because this SRAM memory cell array is used as a cache memory, the number of data entries can thus be increased and the result is that the cash hit rate can be increased.

Furthermore, when the word lines $W_1 \sim W_n$ of the SRAM memory cell array 12 are maintained in the non-active state, during writing operations applied to the DRAM memory cell array, but also during reading operation when reading is performed from the DRAM memory cell array 1, because a configuration can be created in which transmission to the cache memory is not carried out, the advantage is that the extent of the freedom available for the application of the cache system is increased.

Figure 11 is a block diagram showing the construction of a simple cache system using the DRAM element shown in Figure 9.

In Figure 11, the construction of a main memory 30 creates a 1 M byte configuration comprising 8 DRAM elements 31 in a 1 M x 1 construction. The difference between the memory system shown in Figure 11 and the memory system shown in Figure 7 is that the number of the word lines in the SRAM memory cell array 12 and the number of the block segments (set number) of the DRAM element 31 is correspondingly increased in TAG 25 and comparator 26, and the cache hit signal CH and the way address signal WA, which are output from the comparator 26, are input to the DRAM element 31. Here, the way signal has 2 bits.

The reference provided in Figure 6A ~ C and Figure 12 shows operation waveform diagrams used to explain the operation of the conventional simple cache system with the operation of the simple cache system shown in Figure 11.

TAG 25 holds row addresses corresponding to the rows selected in the most recent cycle for each block as an address set using multiple sets of caches. Because in this case, the way address signal can be considered as a 2-bit signal, 4 sets of row addresses are held. Accordingly,

16 address sets will be stored in the TAG 25. In addition, sets of addresses that are often used are held in a fixed manner in the TAG 25.

First, address signal corresponding to data which is requested by the CPU 24 is generated by the address generator 23. The comparator 26 compares the address set stored in the TAG 25 to the multiple bits (2 bits in the example indicated in Figure 9) corresponding to the block segments out of the row address signal RA and column address signal CA of 10 bits in the address signal. After that, if both items coincide, a cache hit is created and the comparator 26 will issue a high-level cache hit signal CH and a way address signal WA for the hit block. A state machine 27 toggles in response to this cache hit signal CH low address strobe signal RAS so that it is maintained on the low level with the column address strobe signal CAS. In response to that, an address multiplexer 22 applies 10-bit column address signal CA to the DRAM element 31 (see Figure 12). At this time, because control is exercised in the DRAM element 31 with the cache hit signal CH as shown in Figure 9, the column address signal CA will not be furnished to the block decoder 13.

[page 6]

Accordingly, a separated state is maintained for the DRAM memory cell array 1 and for the SRAM memory cell array 12. Also, the data in each of the bit line pairs SBL, SBL is read according to the way address signal WA. Further, a conductive state is maintained with a column decoder 60 by an I/O switch 50 in response to the column address signal CA. Because of that, data present in the SRAM memory 120 is output in response to the column address signal CA and the way address signal WA through the I/O band I/O, I/O and the output buffer 9. In the case of such a hit, output data can be obtained at a high speed with the access time t_{CAC} as if it were the page mode.

On the other hand, when the address signal generated from the address generate 23 does not coincide with the used address set which is held in the TAG 25, a cache miss is created, which is why the comparator 26 will not generate high-level cache hit signal CH.

In this case, the state machine 27 performs an ordinary read cycle of the RAS and CAS control signal, and the address multiplexer 22 sequentially supplies to the DRAM element 31 the row address signal RA and the column address signal CA (see Figure 12). Therefore, because in the case of such a cache miss, output data will be obtained with the low-speed access time t_{RAC} , the state machine 27 will generate the wait signal Wait and the wait state is applied to the CPU 24. In the case of the cache miss, the block data contained in memory cells accessed at this time is transmitted when the block decoder 13 is in the conductive state through the transfer gate 110 from the bit line pairs BL, BL of the DRAM memory cell array 1 in one batch to the block of the SRMA memory 120 selected by the way address signal WA. The content stored in this block of the SRAM memory cell 120 can be rewritten in this manner. Also, a new address related to the way address signal WA of the corresponding block will be held in the TAG 25.

Therefore, because the data of a plurality of blocks is held in the SRAM memory cell 12 which is used as a cache memory in the simple cache system which uses the DRAM element shown in Figure 9, the cache hit rate is increased, which makes it possible to increase the number of entries for data sent to TAG 25.

Also, because the DRAM memory cell arrays is accessed here in the case of a cache miss, although transmission of data to the cache memory was performed from the SRAM memory cell array in the indicated example, this transmission can be also prohibited when a non-selective state is created for all the word lines of the SRAM memory cell array. At the same time, there is an option either to select or not to select the transmission to the SRAM memory array cell also when writing operations are applied to the DRAM memory cell array. Furthermore, the embodiment shown in Figure 11 corresponds to a 4-way set of an associative cache.

(Problems to Be Solved By This Invention)

Nevertheless, when a cache hit was achieved according to the simple cache systems described above, the way address signal WA, which was selected from the address signal to access the SRAM memory cell array 12 as a cache memory, was output after a comparison was carried out with the comparator 26. Therefore, because the way address signal WA was supplied to the DRAM element 31, the operation of the word lines of the SRAM memory cell array 12 was delayed, and while the device can be used as a cache memory having a high-speed SRAM memory cell array 12, the disadvantage of the device is that high-speed access time operations cannot be conducted during a hit.

In order to solve the problems mentioned above, the purpose of this invention is to provide a semiconductor memory device having a built-in cache memory, which makes it possible to create the configuration of a cache system in which high-speed access time operations can be conducted during a hit.

(Means to Solve Problems)

The semiconductor device of this invention is equipped with a first memory cell array, a row selection means, a column selection means, a block selection signal input means, a block selection means, a second memory cell array, a region selection signal input means, a region selection means, a data transmission means, a first selection means, and a second selection means.

[page 7]

The first memory cell array comprises a plurality of memory cells arranged in multiple rows and columns, which are also divided into a plurality of blocks with a plurality of column units. The row selection means is used to select each row in multiple memory cells. The column selection means is used to select each column in multiple memory cells. The block selection

signal input means is used to input a block selection signal. The block selection means is used to select one out of the plurality of blocks in the first memory cell.

In addition, the second memory cell array, which comprises multiple static memory cells arranged in multiple rows and multiple columns, is divided into multiple regions and into multiple column units. The region selection means can be used to select any of the multiple regions of the second memory cell array in response to the region selection signal.

The data transmission means is used to transmit data between the blocks of the first memory cell array, selected with the block selection means, and the region of the second memory cell array, selected with the region selection means.

In addition, the first selection means selects any information corresponding to the plurality of static memory cells contained in each region of the second memory cell array. The second selection means is used to perform selection in response to the region selection signal to select any of the multiple information items selected in each region with the first selection means.

(Operation)

In the semiconductor memory device provided with a built-in cache memory relating to this invention, data blocks can be held on the second memory cell array in multiple rows of the first memory cells, and data blocks can be held in different regions of the second memory cell array simultaneously with multiple sets of data blocks having different rows in the same column of the first memory cell array.

Also, the data blocks of different rows in the same column of the first memory cell array can be arranged in the same row of the second memory cell array. Therefore, by using the cache memory of this second memory cell array, the number of entries for data can be effectively increased, which not only makes it possible to increase the cache hit rate, but also enables high-speed access time operations of the cache memory.

(Embodiment)

The following is an explanation of one embodiment of this invention which uses the enclosed figures.

Figure 1 is a block diagram which shows the construction of a DRAM element according to one embodiment of this invention.

As this embodiment is the same as the DRAM element shown in Figure 9 with the exception of the points described below and the same reference symbols are applied to the corresponding parts, an explanation thereof will be omitted when appropriate.

As shown in the figure, a DRAM memory cell array 1 is divided into multiple blocks on

its address space. This embodiment is divided into 4 blocks BK 1 ~ BK4. Moreover, a SRAM memory cell array 12 is divided into multiple ways in multiple column units. However, it is also possible to use a different number of blocks in the DRAM memory cell array 1 and a different number of ways in the SRAM memory array 12.

Between the DRAM memory cell array 1 and the SRAM memory cell array 12 are arranged a sense amplifier part 4, a block transfer gate part 11, an internal I/O band 41, and a way transfer gate part 42. A block decoder 13, responding to one part of the column address signal CA, (2 bits in the case of this embodiment), instructs the block transfer gate part 11 whether and which block data of the DRAM memory cell array 1 is to be transmitted. The way transfer gate part 42 transmits data transmitted to the internal I/O band 41 to one of the ways of the SRAM memory arrays 12.

[page 8]

The way decoder 14 instructs the way transfer gate part 42 whether and to which way the data of the internal I/O band 41 is to be transmitted in response to the address signal applied through the way address buffer 15.

In the SRAM main memory array 12 is deployed a cache row decoder 43, a cache I/O switch part 44, and a cache column decoder 45. The cache row decoder 43 selects 1 row of the SRAM memory array 12 in response to the cache row address signal acquired from the cache address buffer 46. The cache column decoder part 45 selects 1 column in each of the ways in response to the cache column address applied from the cache address buffer 46. The cache address buffer 46 inputs the column address signal CA applied to the DRAM memory cell array 1 as the cache address signal CCA. A part thereof is applied as a cache row address to the cache row decoder 43, and the other part is applied as a cache column address signal to the cache column decoder 43. Multiple SRAM sense amplifiers 47, corresponding to each of the ways of the SRAM memory cell array 12, are connected to the cache I/O switch part 44 through respective I/O line pairs $I/O_A \sim I/O_D$.

The data contained in the SRAM memory cell array 12, selected for each way with the cache row decoder 43 and with the cache column decoder part 45, is detected and amplified by each corresponding SRAM sense amplifier 47. A way selector 48 selects one data item from the data applied via a plurality of SRAM sense amplifiers 47 in response to the way address signal WA applied from the way address buffer 15, and outputs through an output buffer 9b cache output data D_{OUT} to an external part. When data applied to the input buffer 10b as cache input data D_{IN} is written to 1 of the memory cells of the SRAM memory cell array 12, the opposite path to that described above is used.

Figure 1 shows the status when data A_1 , B_1 , C_1 and D_1 in each of the rows of the block BK1 of the DRAM memory cell array 1 is transmitted to the same row of each way A, B, C and of the SRAM memory cell array 12.

Figure 2 is a diagram showing a detailed construction of one part of the configuration shown in Figure 1.

In each of the blocks BK1 ~ BK4 of the DRAM memory cell array 1, the sense amplifier part 4 and the block transfer gate part 11, the sense amplifier part 4 and the block transfer part 11 comprise n sense amplifier parts 4 and n block transfer gates 110, which are provided with corresponding n bit lines $BL_1 \sim BL_n$. Also, the internal I/O band 41 comprises n I/O lines $I/O_1 \sim I/O_n$. The bit lines $BL_1 \sim BL_n$ in each block are connected to respective corresponding I/O line pairs $I/O_1 \sim I/O_n$ through the sense amplifier 40 and block transfer gate 110.

On the other hand, the SRAM memory cell array 12 is divided into 4 ways and each way comprises a SRAM memory cell 120 with n columns, that is to say n bit line pairs $SBL_1 \sim SBL_n$. Each of the ways comprises a way transfer gate part 42, and way transfer gate 420 with corresponding n bit line pairs $SBL_1 \sim SBL_n$. In each of the ways, respective n bit line pairs $SBL_1 \sim SBL_n$ are connected through the way transfer gate 420 to the corresponding I/O line pairs $I/O_1 \sim I/O_n$. The cache I/O switch part 44 comprises cache I/O switches 440 corresponding to respective bit line pairs $SBL_1 \sim SBL_n$ of the SRAM memory cell array 12, and 4 respective corresponding I/O lines $I/O_A \sim I/O_D$. The n bit line pairs $SBL_1 \sim SBL_n$ belonging to each of the ways are connected respectively through the cache I/O switch 440 to the I/O line corresponding to the way. For example, the bit line pairs SBL_1, SBL_n belonging to the way C, are all connected to the I/O line pair I/O_C .

[page 9]

In addition, a cache column decoder part 45 is deployed for each of the ways. The cache column decoder part 45 of each of the ways comprises cache column decoders 450 corresponding to each column. Each of the cache column decoders 450 is connected to the MOS transistor gate of the corresponding cache I/O switch 440.

Figure 3 is a block diagram showing a simple cache system using the DRAM element of Figure 1.

As shown in Figure 3, the configuration of a main memory 30 comprises 1 M byte having 8 DRAM element 31 provided with the 1 M x 1 construction. Unlike in the memory system shown in Figure 11, in the memory system shown in Figure 3, 10-bit address signal corresponding to said column address signal multiplexed by the multiplexer 22 is input to the DRAM elements 31 as cache address signal, replacing the cache hit signal CH output from the comparator 26. Another difference is that the data selector signal, generated by the state machine 27, corresponding to the cache hit signal CH, is input to a data selector 51. The data selector 51 selects and outputs DRAM data DD, applied from the DRAM element 31 in response to the data selector signal DS, or cache data CD.

The operation of the simple cache system shown in Figure 3 will now be explained while referring to the operation of the waveform diagram shown in Figure 4.

TAG 25 holds address sets which are used by a plurality of caches for row addresses corresponding to the rows selected in the most recent cycle for each individual block. In this case, because the way address signal WA can be considered as a 2-bit signal, 4 sets of row address are held. Therefore, when 4 blocks are used, 16 address sets will be stored in TAG 25. In addition, addresses that are often used are held in a fixed manner in the TAG 25.

First, address signal corresponding to the data which is requested by the CPU 24 is generated by the address generator 23.

The comparator 26 compares the address set held in the TAG 25 to multiple bits (2 bits in the example indicated in Figure 3) corresponding to a block segment out of the column address signal CA and the row address signal RA, with 10 bits out of the 20-bit address signal. After that, if there is a coincidence between both items, a cache hit is created, and the comparator 26 will generate the cache hit signal CH on the high level and the way address signal WA for the hit block.

Before the comparison of the address signal is carried out by this comparator 26, assuming that a cache hit occurs, 10-bit cache address signal CCA is input to the DRAM element 31 so that SRAM cell reading operation will proceed. Therefore, when a cache hit occurs and the way address signal WA is input, the desired data is output at a high speed through the cache output buffer 9b as cache data CD, and cache memory data is obtained from the data selector 51 with the data select signal DS, which is generated in response to the cache hit signal CH.

Conversely, when the address signal which has been input to the comparator 26 does not coincide with the address set held in the TAG 25, a cache miss is created, and the cache hit signal CH will not be generated by the comparator 26. Because of that, the cache data CD output from the SRAM memory cell will be ignored. In this case, the state machine 27 performs RAS and CAS signal control in the ordinary read cycle, and the address multiplexer 22 supplies to the DRAM element 31 sequentially the row address signal RA and the column address signal CA (see Figure 4). Therefore, because output data is obtained with the low-speed access time T_{RAC} in the case of such a cache system, wait signal Wait is generated by the state machine 27 and the CPU 24 is brought into the standby state.

[page 10]

In the case of a cache miss, the block data contained in the memory cell accessed at that time is transferred to the I/O line pairs $I/O_1 \sim I/O_n$ of the internal I/O band 41 through the block transfer gate 110 when a conductive state is created by the block data 13. Also, this data is transferred to a suitable SRAM memory cell array 12 through the way transfer gate which is selected by the way address signal WA, and the content stored in the SRAM memory cell 120 in the row

selected by the cache row decoder 43 is rewritten. In addition, a new address set accessed at this time is held in the TAG 25, which relates to the way corresponding to this data block.

As was explained above, the present embodiment makes it possible to increase the number of data entries to the TAG 25 because data corresponding to a plurality of blocks can be held in the SRAM memory cell array 120 as in cache memory. The result is that the probability of a hit is increased, while another effect is that the cache memory can be accessed with an access time at a high speed.

(Effect of the Invention)

As was explained above, since the present invention makes it possible to hold a great number of data blocks of the first memory cell array in the second memory cell array without increasing unnecessarily the block size, the number of data entries can thus be effectively increased.

Moreover, because data blocks of different rows relating to one column of the first memory cell array is stored in the same row on the second memory cell array, any of the regions can be selected after the information has been read from each region of the second memory cell array. Therefore, since the second memory cell array can be used as a cache memory, access is enabled at an extremely high speed when there is a cache hit. Accordingly, when the semiconductor memory device of this invention is used, this makes it possible to create the configuration of a simple set-associative cache system enabling cache hit operations at a high speed.

4. Brief Explanation of Figures

Figure one is a construction block diagram of a semiconductor memory device according to one embodiment of this invention, Figure 2 is a block diagram showing the details of the construction of one part of the semiconductor memory device shown in Figure 1, Figure 3 is a block diagram showing the construction of a simple set-associative cache system which utilizes the semiconductor memory device shown in Figure 1, Figure 4 is an operation waveform diagram explaining the operation of the simple cache system of Figure 3, Figure 5 is a block diagram showing the construction of a conventional DRAM element, Figure 6A is an operation waveform diagram of an ordinary read cycle of a DRAM element according to prior art, Figure 6B is an operation waveform diagram of the page mode cycle of a DRAM element according to prior art, Figure 6C is an operation waveform diagram of the static column mode of a DRAM element according to prior art, Figure 7 is a block diagram showing the construction of a simple cache system utilizing the DRAM element shown in Figure 5, Figure 8 is an operation waveform diagram of the simple cache system of Figure 7, Figure 9 is a block diagram showing the construction of a DRAM element provided with a built-in cache memory, Figure 10 is a block diagram showing the detailed construction of one part of the DRAM element of Figure 9, Figure 11 is a block diagram showing the construction of a simple cache system utilizing the DRAM element of Figure 9, and Figure 12 is an operation waveform diagram of the simple cache system of Figure 11.

In these figures, 1 is a DRAM memory cell array, 2 is a word driver, 3 is a row decoder part, 4 is a sense amplifier part, 5 is an I/O switch part, 6 is a column decoder part, 7 is a row address buffer, 8 is a column address buffer, 9a, 9b are output buffers, 10a, 10b are input buffers, 11 is a block transfer gate part, 12 is a SRAM memory cell array, 13 is a block decoder, 14 is a way decoder, 15 is a way address buffer, 41 is a built-in I/O belt, 42 is a way transfer gate part, 43 is a cache row decoder, 44 is a cache I/O switch part, 45 is a cache column decoder part, 46 is a cache address buffer, 47 is a sense amplifier for SRAM, 48 is a way selector, BL, BL is a pair of bit lines of a DRAM memory cell array, and SBL, SLB is a pair of bit lines of a SRAM memory cell array.

In addition, the same symbols indicate the same or corresponding parts in the figures.

[page 11]

Figure 1

- 1 DRAM memory cell array
- 2 word driver
- 3 row decoder part
- 4 sense amplifier
- 5 I.O switch part
- 6 normal column decoder
- 7 row address buffer
- 8 column address buffer
- 9a output buffer
- 9b output buffer B cache D_{OUT}
- 10a input buffer
- 10b input buffer B cache D_{IN}
- 11 block transfer gate part
- 12 SRAM memory cell array
- 13 block decoder
- 14 way decoder
- 15 way address buffer
- 41 built-in I/O belt
- 42 way transfer gate part
- 43 cache column decoder
- 44 cache I/O switch part
- 45 cache column decoder part
- 47 SRAM sense amplifier
- 48 way selector

Figure 2

- (1) way decoder output
- (2) block decoder output
- (3) the result of blocks BK3 and BK4
- (4) DRAM memory cell array
- (5) the result of way A and B
- 450 cache column decoder
- 450 cache column decoder
- 440 cache I/O switch
- 120 SRAM memory cell
- 120 SRAM memory cell
- 120 SRAM memory cell
- 420 way transfer gate
- 110 block transfer gate
- 45: cache column decoder part
- 44: cache I/O switch part
- 12: SRAM memory cell array
- 42: way transfer gate part
- 41: built-in I/O belt
- 11: block transfer gate part
- 1: DRAM memory cell array

[page 12]

Figure 3

- 22 address multiplexer
- 23 address generator
- 25 latch (TAG)
- 26 comparator
- 27 state machine
- 30: main memory (1 B byte)
- 51: data selector

Figure 4

- (1) system block
- (2) address (20)
- (3) cache Dout
- (4) high impedance

Figure 5

- 1 memory cell array
- 2 word driver
- 3 row decoder part
- 4 sense amplifier
- 5 I/O switch part
- 6 column decoder part
- 7 row address buffer
- 8 column address buffer
- 9 output buffer D_{OUT}
- 10 input buffer D_{IN}

Figure 6A

- (1) ordinary cycle
- (2) valid
- (3) page mode cycle

Figure 6B

- (3) page mode cycle
- (4) valid
- (5) valid

Figure 6C

- (6) static column mode
- (7) valid
- (8) valid

[page 13]

Figure 7

- 20: main memory (1 M byte)
- 22: address multiplexer
- 23: address generator
- 25: latch (TAG)
- 26: comparator
- 27: state machine

Figure 8

- (1) system clock
- (2) address (20)

Figure 9

- 1 DRAM memory arrays B1 ~ B4
- 2 word driver
- 3 row decoder part
- 4 sense amplifier part
- 5 I/O switch part
- 5 block decoder
- 6 column decoder part
- 7 row address buffer
- 8 column address buffer
- 9 output buffer D_{OUT}
- 10 input buffer D_{IN}
- 11 transfer gate part
- 12 SRAM memory cell array
- 13 block decoder
- 14 way decoder
- 15 way address buffer

Figure 10

- (A) block boundary
- 1 DRAM memory array
- 12: SRAM memory cell array
- 13 block decoder
- 50 I/O switch
- 60 column decoder
- 60 column decoder
- 110 transfer gate
- 120 SRAM memory cell

[page 14]

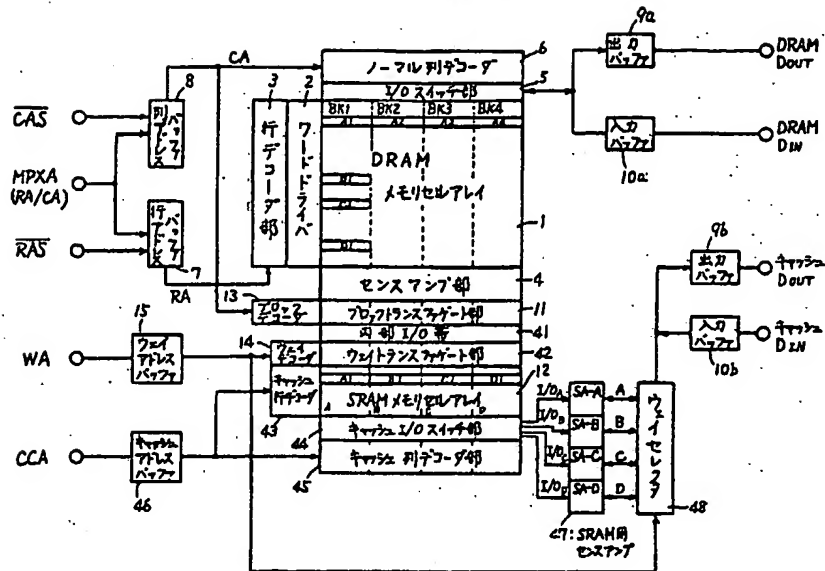
Figure 11

- 22 address multiplexer
- 23 address generator
- 25 latch (TAG)
- 26 comparator
- 27 state system
- 30 main memory (1 M byte)

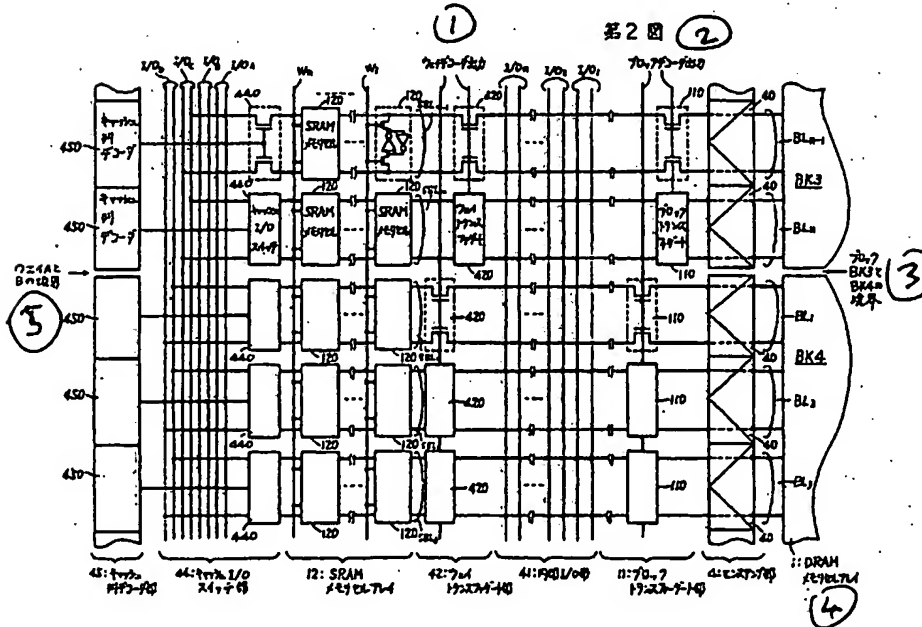
Figure 12

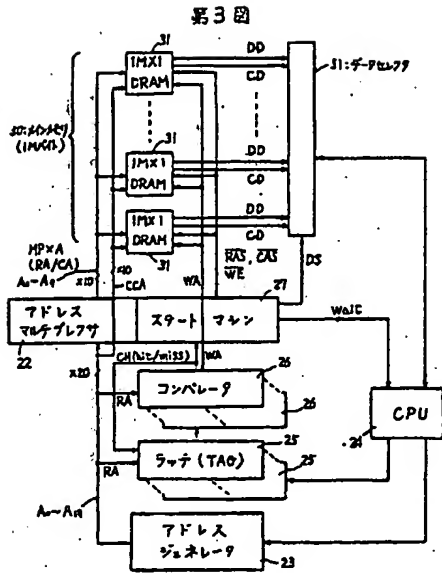
- (1) system clock
- (2) address (20)

第1圖

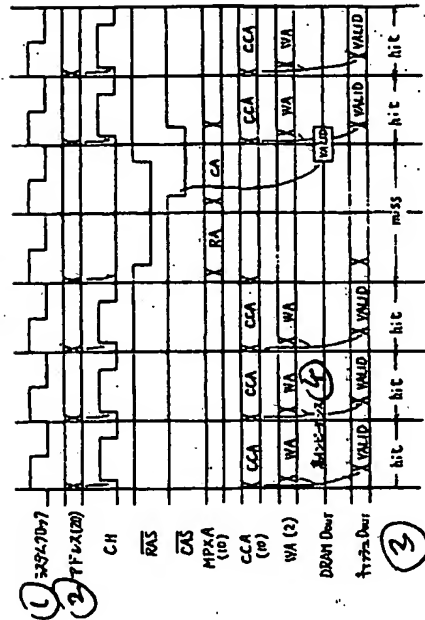


第2圖 (2)

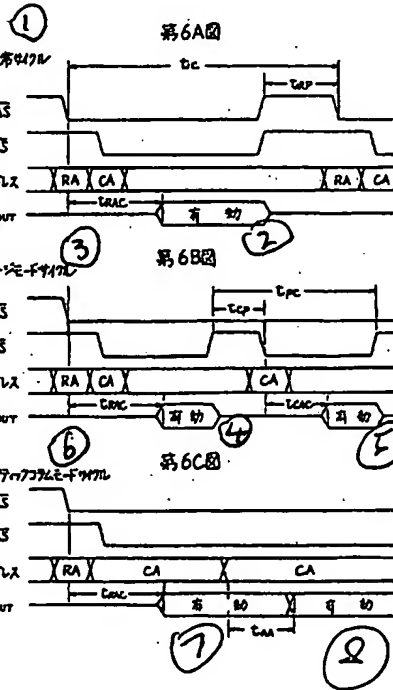
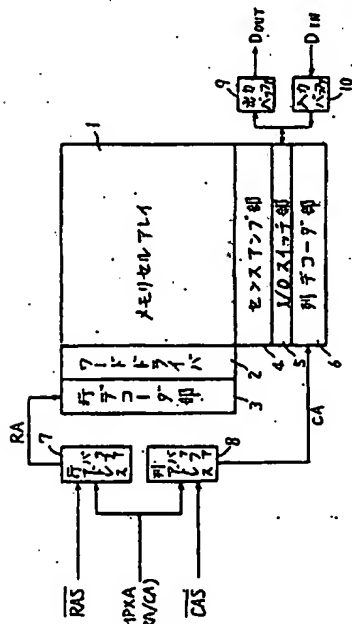


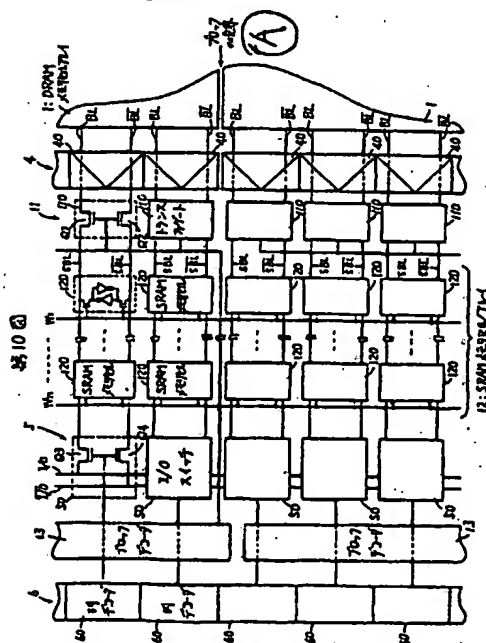
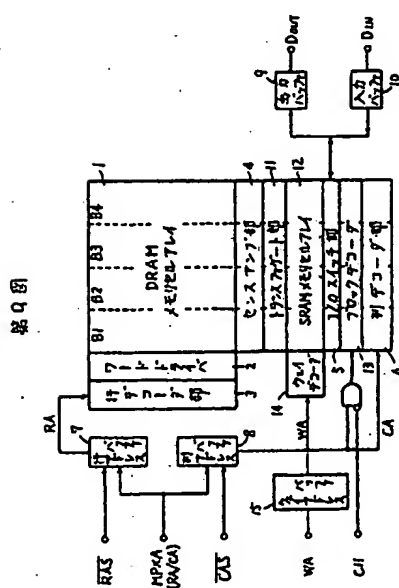
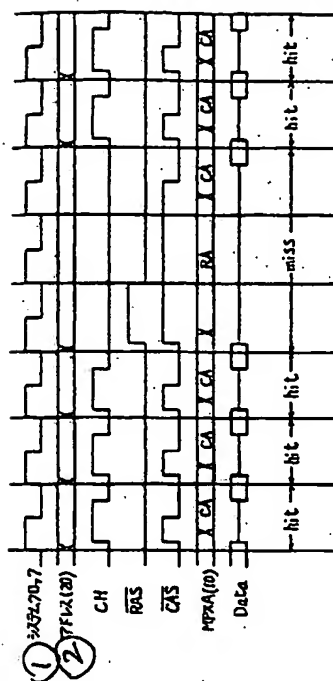
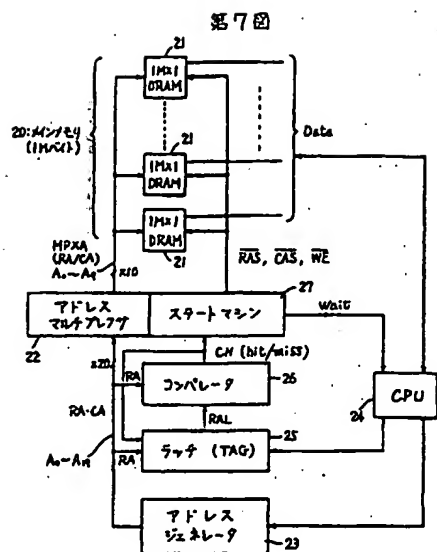


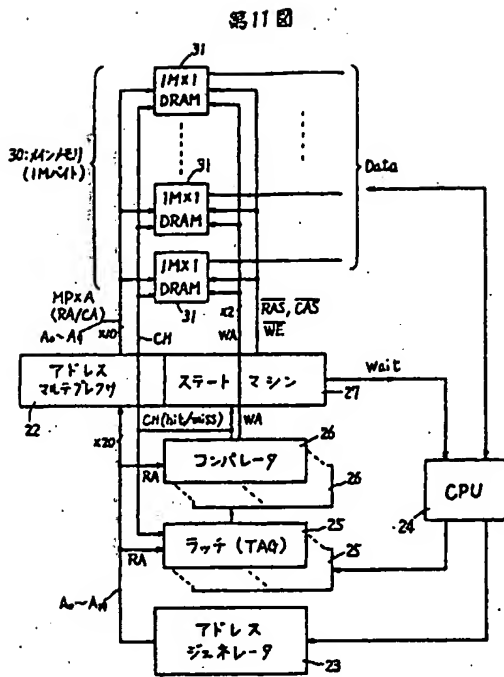
第4図



第5図







第12図

